

We claim:

1 1. Apparatus for providing a visualized hierarchical
2 display of categorized event data, said data being a
3 collection of records, wherein each record is
4 associated with an occurrence of a corresponding event
5 and comprises a plurality of attribute/value pairs
6 characterizing the event or an individual user
7 associated with the event, the apparatus comprising:

8 a processor;
9 a memory connected to the processor and storing
10 computer executable instructions therein;

11 circuitry, connected to the processor, for
12 accessing a plurality of data records, residing on a
13 data storage medium, that collectively forms a dataset
14 representing a collection of events and for applying
15 the data records from the medium to the processor; and
16 a display operative in conjunction with the
17 processor;

18 wherein the processor, in response to execution of
19 the stored instructions:

20 classifies the data records, based on the
21 attribute/value pairs associated with each such record,
22 into a plurality of mutually exclusive first clusters;

23 determines a measure of similarity between
24 each pair of said first clusters so as to yield a
25 plurality of similarity measures for the first clusters
26 representing the dataset; and

27 forms, based on the similarity measures, a
28 multi-level hierarchical cluster organization such that
29 said first clusters are situated, as leaf nodes, at a

30 lowest level of a hierarchy with second clusters being
31 situated, as cluster group nodes, at successively
32 higher levels of the hierarchy and formed as a result
33 of selectively and iteratively combining clusters that
34 are sufficiently similar to each other so as to form
35 combined clusters in order to define a nodal set
36 wherein each of the combined clusters replaces the
37 clusters so combined to form said each combined
38 cluster; and

39 visually renders the hierarchical
40 organization on the display.

1 2. The apparatus in claim 1 wherein the processor, in
2 response to execution of the stored instructions:

3 summarizes each of the first clusters into a
4 corresponding first segment so as to define a plurality
5 of first segments such that each of said first segments
6 contains records, from within its associated one of the
7 first clusters, that exhibit similar behavior and
8 similar properties;

9 determines the similarity measures between each
10 pair of said first segments so as to yield a plurality
11 of similarity measures; and

12 forms the multi-level hierarchical organization,
13 through agglomerative clustering, of the first
14 segments.

1 3. The apparatus in claim 2 wherein the processor, in
2 response to execution of the stored instructions, forms
3 a root node that represents the entire collection and
4 is situated at a highest level of the hierarchy.

1 4. The apparatus in claim 3 wherein the processor, in
2 response to execution of the stored instructions,
3 performs agglomerative clustering by:
4 (a) determining a measure of distance between each
5 pair of members in the nodal set, the nodal set
6 initially being defined as having all of said members,
7 as child nodes in the hierarchy, and,
8 (b) for each pair of said members having nearest
9 distances therebetween, combining said pair of members
10 to yield a parent node at a higher level of the
11 hierarchy, wherein the parent node replaces the pair of
12 said members in the nodal set; and
13 (c) iteratively repeating operations (a) and (b) until
14 the root node is formed that represents all the members
15 in the collection.

1 5. The apparatus in claim 4 wherein the processor, in
2 response to execution of the stored instructions,
3 reduces a level of the hierarchy by successively
4 merging, based on nearest similarity measures, two
5 linked nodes at adjacent levels in the hierarchy so as
6 to form a single substitute node having a group of
7 segments associated with the two nodes that have been
8 merged.

1 6. The apparatus in claim 4 wherein the processor, in
2 response to execution of the stored instructions:
3 accepts a user-selection of a segment in the
4 hierarchy so as to define a first selected segment;

5 scores each of the attribute/value pair associated
6 with the first selected segment as to how well each of
7 said attribute/value pairs associated with the first
8 selected segment characterizes the first selected
9 segment;

10 rank orders the attribute/value pairs within the
11 first selected segment so as to define a first rank
12 order; and

13 visually displays each one of a plurality of the
14 attribute/value pairs within the first selected segment
15 in said first rank order along with an indication
16 representative of a magnitude of the score of said one
17 of the plurality of said attribute/value pairs within
18 the first selected segment.

1 7. The apparatus in claim 6 wherein the indication is
2 graphical.

1 8. The apparatus in claim 7 wherein each of the
2 records reflects a user who visits a predefined web
3 site with the attributes in the record reflecting
4 information regarding a transaction in which the user
5 has engaged with the web site or characteristic
6 information, regarding the user, which the user has
7 furnished to the web site.

1 9. The apparatus in claim 7 wherein the processor, in
2 response to execution of the stored instructions,
3 determines the score of each of the attribute/value
4 pairs on a discriminative basis.

1 10. The apparatus in claim 7 wherein the processor, in
2 response to execution of the stored instructions:

3 generates a graphical user interface on the
4 display; and

5 selectively expands or contracts the displayed
6 hierarchy based on input commands based on user input
7 from an individual interacting with the apparatus
8 through the graphical user interface.

1 11. The apparatus in claim 4 wherein the processor, in
2 response to execution of the stored instructions:

3 accepts user-selection of a pair of segments in
4 the hierarchy so as to define first an second selected
5 segments in the hierarchy;

6 scores each of the events associated with the
7 second selected segment as to how well each of the
8 attribute/value pairs associated with the second
9 selected segment characterizes events associated with a
10 first selected segment;

11 rank orders the attribute/value pairs associated
12 with the second selected segment so as to define a
13 second rank order; and

14 visually displays each one of a plurality of the
15 attribute/value pairs associated with the second
16 selected segment in said second rank order along with
17 an indication representative of a magnitude of the
18 score of each one of the plurality of said
19 attribute/value pairs so as to facilitate a visual
20 comparison of the attribute/value pairs of the first
21 and second selected segments and to visually assess
22 whether each of the plurality of said attribute/value

23 pairs associated with the second segment is more likely
24 to be exhibited by the first or second selected
25 segments.

1 12. The apparatus in claim 11 wherein the processor,
2 in response to the stored instructions, determines the
3 score of each of the events associated with the second
4 segment based on corresponding probabilities of said
5 each event occurring or not occurring in all of the
6 segments.

1 13. The apparatus in claim 12 wherein the processor,
2 in response to the stored instructions, ascertains the
3 corresponding probabilities in response to the
4 attribute/value pairs associated with said each event.

1 14. The apparatus in claim 11 wherein the processor,
2 in response to the stored instructions, determines the
3 score of said each of the events associated with the
4 second segment through use of discriminant values.

1 15. The apparatus in claim 11 wherein the indication
2 is graphical.

1 16. The apparatus in claim 15 wherein each of the
2 records reflects a user who visits a predefined web
3 site with the attributes in the record reflecting
4 information regarding a transaction in which the user
5 has engaged with the web site or characteristic
6 information, regarding the user, which the user has
7 furnished to the web site.

1 17. The apparatus in claim 15 wherein the processor,
2 in response to execution of the stored instructions:

3 generates a graphical user interface on the
4 display; and

5 selectively expands or contracts the displayed
6 hierarchy based on input commands based on user input
7 from an individual interacting with the apparatus
8 through the graphical user interface.

1 18. The apparatus in claim 15 wherein the processor,
2 in response to the stored instructions, limits a depth
3 of the hierarchy to a predefined level.

1 19. A method, for use in conjunction with apparatus,
2 for providing a visualized hierarchical display of
3 categorized event data, said data being a collection of
4 records, wherein each record is associated with an
5 occurrence of a corresponding event and comprises a
6 plurality of attribute/value pairs characterizing the
7 event or an individual user associated with the event,
8 the apparatus having: a processor; a memory connected
9 to the processor and storing computer executable
10 instructions therein; circuitry, connected to the
11 processor, for accessing a plurality of data records,
12 residing on a data storage medium, that collectively
13 forms a dataset representing a collection of events and
14 for applying the data records from the medium to the
15 processor; and a display operative in conjunction with
16 the processor; wherein the method comprises the steps

17 performed by the processor, in response to execution of
18 the stored instructions, of:

19 classifying the data records, based on the
20 attribute/value pairs associated with each such record,
21 into a plurality of mutually exclusive first clusters;

22 determining a measure of similarity between each
23 pair of said first clusters so as to yield a plurality
24 of similarity measures for the first clusters
25 representing the dataset; and

26 forming, based on the similarity measures, a
27 multi-level hierarchical cluster organization such that
28 said first clusters are situated, as leaf nodes, at a
29 lowest level of a hierarchy with second clusters being
30 situated, as cluster group nodes, at successively
31 higher levels of the hierarchy and formed as a result
32 of selectively and iteratively combining clusters that
33 are sufficiently similar to each other so as to form
34 combined clusters in order to define a nodal set
35 wherein each combined cluster replaces the clusters so
36 combined to form said combined clusters; and

37 visually renders the hierarchical
38 organization on the display.

1 20. The method in claim 19 further comprising the
2 steps of:

3 summarizing each of the first clusters into a
4 corresponding first segment so as to define a plurality
5 of first segments such that each of said first segments
6 contains records, from within its associated one of the
7 first clusters, that exhibit similar behavior and
8 similar properties;

9 determining the similarity measures between each
10 pair of said first segments so as to yield a plurality
11 of similarity measures; and

12 forming the multi-level hierarchical organization,
13 through agglomerative clustering, of the first
14 segments.

1 21. The method in claim 20 further comprising the step
2 of forming a root node that represents the entire
3 collection and is situated at a highest level of the
4 hierarchy.

1 22. The method in claim 21 wherein agglomerative
2 clustering is performed by:

3 (a) determining a measure of distance between each
4 pair of members in a nodal set, the nodal set initially
5 being defined as having all of said members, as child
6 nodes in the hierarchy, and,

7 (b) for each pair of said members having nearest
8 distances therebetween, combining said pair of members
9 to yield a parent node at a higher level of the
10 hierarchy, wherein the parent node replaces the pair of
11 said members in the nodal set; and

12 (c) iteratively repeating operations (a) and (b) until
13 the root node is formed that represents all the members
14 in the collection.

1 23. The method in claim 22 further comprising the step
2 of reducing a level of the hierarchy by successively
3 merging, based on nearest similarity measures, two
4 linked nodes at adjacent levels in the hierarchy so as

5 to form a single substitute node having a group of
6 segments associated with the two nodes that have been
7 merged.

1 24. The method in claim 22 further comprising the
2 steps of:

3 accepting a user-selection of a segment in the
4 hierarchy so as to define a first selected segment;
5 scoring each of the attribute/value pairs within
6 the first selected segment as to how well each of said
7 attribute/value pairs associated with the first
8 selected segment characterizes the first selected
9 segment;

10 rank ordering the attribute/value pairs within the
11 first selected segment so as to define a first rank
12 order; and

13 visually displaying each one of a plurality of the
14 attribute/value pairs within the first selected segment
15 in said first rank order along with an indication
16 representative of a magnitude of the score of said one
17 of the plurality of said attribute/value pairs within
18 the first selected segment.

1 25. The method in claim 24 wherein the indication is
2 graphical.

1 26. The method in claim 25 wherein each of the records
2 reflects a user who visits a predefined web site with
3 the attributes in the record reflecting information
4 regarding a transaction in which the user has engaged
5 with the web site or characteristic information,

6 regarding the user, which the user has furnished to the
7 web site.

1 27. The method in claim 24 further comprising the step
2 of determining the score of each of the attribute/value
3 pairs on a discriminative basis.

1 28. The method in claim 25 further comprising the
2 steps of:

3 generating a graphical user interface on the
4 display; and

5 selectively expanding or contracting the displayed
6 hierarchy based on input commands based on user input
7 from an individual interacting with the apparatus
8 through the graphical user interface.

1 29. The method in claim 22 further comprising the
2 steps of:

3 accepting a user-selection of a pair of segments
4 in the hierarchy so as to define first and second
5 selected segments in the hierarchy;

6 scoring each of the attribute/value pairs
7 associated with the second selected segment as to how
8 well each of said attribute/value pairs associated with
9 the second selected segment characterizes events
10 associated with a first selected segment;

11 rank ordering the attribute/value pairs associated
12 with the second selected segment so as to define a
13 second rank order; and

14 visually displaying each one of a plurality of the
15 attribute/value pairs associated with the second

16 selected segment in said second rank order along with
17 an indication representative of a magnitude of the
18 score of each one of the plurality of said
19 attribute/value pairs, so as to facilitate a visual
20 comparison of the attribute/value pairs of the first
21 and second selected segments and to visually assess
22 whether each of the plurality of said attribute/value
23 pairs associated with the second segment is more likely
24 to be exhibited by the first or second selected
25 segments.

1 30. The method in claim 29 wherein the scoring step
2 comprises the step of determining the score of each of
3 the events associated with the second segment based in
4 corresponding probabilities of said each event
5 occurring or not occurring in all of the segments.

1 31. The method in claim 30 wherein the score
2 determining step comprises the step of ascertaining the
3 corresponding probabilities in response to the
4 attribute/value pairs associated with said each event.

1 32. The method in claim 29 wherein the scoring step
2 comprises the step of determining the score of said
3 each of the events in the second segment through use of
4 discriminant values.

1 33. The method in claim 29 wherein the indication is
2 graphical.

1 34. The method in claim 33 wherein each of the records
2 reflects a user who visits a predefined web site with
3 the attributes in the record reflecting information
4 regarding a transaction in which the user has engaged
5 with the web site or characteristic information,
6 regarding the user, which the user has furnished to the
7 web site.

1 35. The method in claim 33 further comprising the
2 steps of:

3 generating a graphical user interface on the
4 display; and

5 selectively expanding or contracting the displayed
6 hierarchy based on input commands based on user input
7 from an individual interacting with the apparatus
8 through the graphical user interface.

1 36. The method in claim 35 further comprising the step
2 of limiting a depth of the hierarchy to a predefined
3 level.

1 37. A computer readable medium having computer
2 executable instructions stored therein, said
3 instructions being executed by a computer, for
4 performing the steps in claim 19.

1 38. Apparatus for providing a visualized hierarchical
2 display of categorized event data, said data being a
3 collection of records, wherein each record is
4 associated with an occurrence of a corresponding event
5 and comprises a plurality of attribute/value pairs

6 characterizing the event or an individual user
7 associated with the event, the apparatus comprising:
8 a processor;
9 a memory connected to the processor and storing
10 computer executable instructions therein;
11 circuitry, connected to the processor, for
12 accessing a plurality of data records, residing on a
13 data storage medium, that collectively forms a dataset
14 representing a collection of events and for applying
15 the data records to the processor; and
16 a display operative in conjunction with the
17 processor;
18 wherein the processor, in response to execution of
19 the stored instructions:
20 automatically classifies the data records,
21 based on the attribute/value pairs associated with each
22 such record, into a plurality of mutually exclusive
23 clusters;
24 determines a measure of similarity between
25 each pair of said clusters so as to yield a plurality
26 of similarity measures for the first clusters
27 representing the dataset; and
28 visually renders each one of said pairs of
29 clusters on the display along with a visual indication
30 of a corresponding one of the similarity measures which
31 is associated with said each pair of said clusters.

1 39. The apparatus in claim 38 wherein the processor,
2 in response to execution of the stored instructions:
3 establishes a similarity threshold; and

4 displays the indication of the determined
5 similarity measure for said each pair of clusters if
6 the determined similarity measure exceeds the
7 similarity threshold.

1 40. The apparatus in claim 38 wherein the visual
2 indication comprises thickness of a displayed arc that
3 connects the first and second clusters, a color of the
4 arc or other visual characteristic of the arc.

1 41. The apparatus in claim 40 wherein the processor,
2 in response to execution of the stored instructions:
3 establishes a similarity threshold; and
4 displays the indication of the determined
5 similarity measure for said each pair of clusters if
6 the determined similarity measure exceeds the
7 similarity threshold.

1 42. The apparatus in claim 38 wherein the processor,
2 in response to execution of the stored instructions:
3 receives an instruction to de-emphasize a
4 particular cluster; and
5 in response to the instruction to de-emphasize a
6 cluster, de-emphasizes the visual indication for the
7 particular cluster.

1 43. The apparatus in claim 38 wherein the processor,
2 in response to execution of the stored instructions,
3 receives a user-specified level for the similarity
4 threshold.

1 44. The apparatus in claim 43 wherein the processor,
2 in response to execution of the stored instructions,
3 displays a slider through which the user can set the
4 similarity threshold.

1 45. The apparatus in claim 43 wherein the visual
2 indication is a displayed arc that connects the first
3 and second clusters and the processor, in response to
4 execution of the stored instructions, displays, with
5 the slider set to one end position, either no or a
6 minimum number of arcs between corresponding ones of
7 the clusters and, with the slider set to another end
8 position, all pair-wise connections.

1 46. The apparatus in claim 44 wherein the processor,
2 in response to execution of the stored instructions,
3 displays the slider either horizontally or vertically.

1 47. The apparatus in claim 43 wherein the processor,
2 in response to execution of the stored instructions,
3 adjusts the displayed indication of the similarity
4 measure for said each cluster to reflect a change in
5 the user-specified similarity threshold.

1 48. The apparatus in claim 38 wherein the hierarchical
2 display is visually arranged as a spring model wherein
3 apparent attraction force between said each pair of the
4 clusters is responsive to the similarity measure for
5 said each pair of clusters.

1 49. The apparatus in claim 38 wherein the processor,
2 in response to execution of the stored instructions:
3 receives a user-supplied instruction to split a
4 particular displayed cluster; and
5 in response to the user-supplied instruction,
6 displays a pair of clusters for the particular
7 displayed combined cluster.

1 50. The apparatus in claim 49 wherein the processor,
2 in response to execution of the stored instructions,
3 displays a slider wherein user movement of the slider
4 specifies a corresponding similarity measure, for the
5 pair of clusters, sufficient to split the particular
6 displayed combined cluster into said pair of clusters.

1 51. The apparatus in claim 50 wherein the processor,
2 in response to execution of the stored instructions,
3 displays an animation of splitting the particular
4 displayed cluster into said pair of clusters.

1 52. The apparatus in claim 49 the particular displayed
2 cluster is a displayed cluster that resulted from a
3 most recent combination of a pair of clusters.

1 53. A method, for use in conjunction with apparatus,
2 for providing a visualized hierarchical display of
3 categorized event data, said data being a collection of
4 records, wherein each record is associated with an
5 occurrence of a corresponding event and comprises a
6 plurality of attribute/value pairs characterizing the
7 event or an individual user associated with the event,

8 the apparatus having: a processor; a memory connected
9 to the processor and storing computer executable
10 instructions therein; circuitry, connected to the
11 processor, for accessing a plurality of data records,
12 residing on a data storage medium, that collectively
13 forms a dataset representing a collection of events and
14 for applying the data records to the processor; and a
15 display operative in conjunction with the processor;
16 the method comprising the steps, performed by the
17 processor, in response to execution of the stored
18 instructions, of:

19 automatically classifying the data records, based
20 on the attribute/value pairs associated with each such
21 record, into a plurality of mutually exclusive
22 clusters;

23 determining a measure of similarity between each
24 pair of said clusters so as to yield a plurality of
25 similarity measures for the first clusters representing
26 the dataset; and

27 visually rendering each one of said pairs of
28 clusters on the display along with a visual indication
29 of a corresponding one of the similarity measures which
30 is associated with said each pair of said clusters.

1 54. The method in claim 53 further comprising the
2 steps of:

3 establishing a similarity threshold; and

4 displaying the indication of the determined
5 similarity measure for said each pair of clusters if
6 the determined similarity measure exceeds the
7 similarity threshold.

1 55. The method in claim 53 wherein the visual
2 indication comprises thickness of a displayed arc that
3 connects the first and second clusters, a color of the
4 arc or other visual characteristic of the arc.

1 56. The method in claim 55 further comprising the
2 steps of:

3 establishing a similarity threshold; and
4 displaying the indication of the determined
5 similarity measure for said each pair of clusters if
6 the determined similarity measure exceeds the
7 similarity threshold.

1 57. The method in claim 53 further comprising the
2 steps of:

3 receiving an instruction to de-emphasize a
4 particular cluster; and
5 in response to the instruction to de-emphasize a
6 cluster, de-emphasizing the visual indication for the
7 particular cluster.

1 58. The method in claim 53 further comprising the step
2 of receiving a user-specified level for the similarity
3 threshold.

1 59. The method in claim 58 further comprising the step
2 of displaying a slider through which the user can set
3 the similarity threshold.

1 60. The method in claim 59, wherein the visual
2 indication is a displayed arc that connects the first
3 and second clusters, comprising the step of displaying
4 the stored instructions, with the slider set to one end
5 position, either no or a minimum number of arcs between
6 corresponding ones of the clusters and, with the slider
7 set to another end position, all pair-wise connections.

1 61. The method in claim 59 further comprising the step
2 of displaying the slider either horizontally or
3 vertically.

1 62. The method in claim 58 further comprising the step
2 of adjusting the displayed indication of the similarity
3 measure for said each cluster to reflect a change in
4 the user-specified similarity threshold.

1 63. The method in claim 53 further comprising the step
2 of visually arranging the hierarchical display arcs as
3 a spring model wherein apparent attractive force
4 between said each pair of the clusters is responsive to
5 the similarity measure for said each pair of clusters.

1 64. The method in claim 53 further comprising the
2 steps of:

3 receiving a user-supplied instruction to split a
4 particular displayed cluster; and

5 in response to the user-supplied instruction,
6 displaying a pair of clusters for the particular
7 displayed combined cluster.

1 65. The method in claim 64 further comprising the step
2 of displaying a slider wherein user movement of the
3 slider specifies a corresponding similarity measure,
4 for the pair of clusters, sufficient to split the
5 particular displayed combined cluster into said pair of
6 clusters.

1 66. The method in claim 65 further comprising the step
2 of displaying an animation of splitting the particular
3 displayed cluster into said pair of clusters.

1 67. The method in claim 64 wherein the particular
2 displayed cluster is a displayed cluster that resulted
3 from a most recent combination of a pair of clusters.

1 68. A computer readable medium having computer
2 executable instructions stored therein, said
3 instructions being executed by a computer, for
4 performing the steps in claim 53.